Regularized Synthetic Control Methods:

Advancing Causal Inference in Time Series Econometrics and Observational Studies

Jörg Breitung[a], **Lennart Bolwin**[a], **Justus Töns**[a]

[a]University of Cologne, Chair of Statistics and Econometrics

---

Abstract

The Synthetic Control (SC) method is a widely used tool for measuring causal treatment effects in observational trials. Typically, the counterfactual of the single treated unit is synthesized using a weighted average of the remaining units in the post-treatment phase. These weights are computed in a data-driven manner and aim to minimize the distance between the treated unit and its counterfactual in the pre-treatment phase. To avoid overfitting the training data and to ensure external validity of the results, the method's developers (Abadie, Diamond, and Hainmueller (ADH)) incorporated the constraint that each weight must be weakly positive, and all weights must up sum to one. Building on the work of [Doudchenko and Imbens, 2016], we propose a generalization that allows for the inclusion of a constant term and negative weights. However, we develop the Regularized Synthetic Controls (REGSC) estimator, an alternative regularization approach that shrinks individual coefficients towards zero and the sum of coefficients towards one. Besides the crucial advantage of a closed form expression, Monte Carlo studies confirm that this regularization method dominates other estimators in data-generating processes with factor structure as was proposed by the inventors of SC. Next, we extend the approach to dynamic contexts and propose a regularized Autoregressive Distributed Lag (ARDL) model for optimal estimation of the counterfactual in time series settings. Again, simulations confirm the new method's potential to enhance the accuracy and robustness of causal effect estimation in time series econometrics and observational studies.

Keywords: *Synthetic Control; Observational Studies; Causal Inference; Regularization, Autoregressive Distributed Lag Models*

List of Acronyms

ADH Abadie, Diamond, and Hainmueller

ARDL Autoregressive Distributed Lag

GDP Gross Domestic Product

DGP Data Generating Process

iid independent and identically distributed

MSFE Mean Squared Forecast Error

MSPE Mean Squared Prediction Error

OLS Ordinary Least Squares

PC Principal Components

REGSC Regularized Synthetic Controls

RSS Residual Sum of Squares

SC Synthetic Control

USA United States of America

VAR Vector Autoregression

VARSC Vector Autoregressive Synthetic Control

Contents

List of Figures

1. Introduction

This is work in progress

SC method is combined with Vector Autoregression (VAR). Method was introduced by ADH. Paper strcuture could be similar to the one by [Doudchenko and Imbens, 2016]

2. Literature Review 2-3 pages

Literature has been clustered. This is work in progress

2.1. Synthetic Control

The SC method was developed by Alberto Abadie and colleagues in a series of influential papers ([Abadie and Gardeazabal, 2003], [Abadie et al., 2010], [Abadie et al., 2015]). The method is designed to estimate the causal effect of a treatment in a setting with a single treatment unit and a number of potential control units. Pre- and post-treatment data are observed for the treatment and control units for the outcome of interest as well as for a set of covariates. The SC-procedure combines aspects of the matching and difference-in-difference literature and can therefore be interpreted as a relative of the causal inference literature introduced by [Rubin, 1974]. Similar to many other microeconometric methods, the objective is to distinguish causation from correlation and to assess the magnitude and significance of treatments in observational case studies.
In their canonical 2003 article, Abadie and Gardeazabal evaluate the causal economic effects of conflict using terrorist conflicts in the Basque Country as a comparative case study. In their specific application example, they find that terrorist conflicts caused the per capita Gross Domestic Product (GDP) of the treatment unit (Basque Country) to decline by about 10% relative to the synthesized control unit.

some more words on other findis and things they did The next appropriate setting for an application of the SC method was the introduction of a large-scale tobacco control program implemented in the state of California in the United States of America (USA) in 1988.

2.2. Overview

[Abadie, 2021] read.
[Athey and Imbens, 2016] read.

2.3. Application

[Born et al., 2019] read.

[Cho, 2020] read.

[Cunningham, 2021] read.

[Funke et al., 2020] read.

2.4. Methodological Background

[Hainmueller et al., 2011] read.

[Abadie and Imbens, 2006] not read.

[Abadie and Imbens, 2002] not read.

[Doudchenko and Imbens, 2016] read.

[Ferman, 2021] read.

[Frangakis and Rubin, 2002] not read.

[Rosenbaum and Rubin, 1983] not read

[Rubin, 1974] not read.

2.5. Extensions/ Developments

[Abadie and L'Hour, 2021] read.

[Amjad et al., 2018] read.

[Ben-Michael et al., 2021] read.

[Ben-Michael et al., 2021] not read.

[Kellogg et al., 2021] not read.

[Kuosmanen et al., 2021] not read.

[Muhlbach and Nielsen, 2019] read.

Developments

[Arkhangelsky et al., 2021] not read

[Athey et al., 2017] not read.

[Brodersen et al., 2015] read.

[von Brzeski et al., 2015] read.

[Hartford et al., 2017] read.

2.6. Testing

[Andrews, 2003] not read.

[Cattaneo et al., 2021] not read.

[Chernozhukov et al., 2019] not read.

[Chernozhukov et al., 2021] not read.

[Firpo and Possebom, 2018] not read.

[Hahn and Shi, 2017] read.

2.7. Time Series Econometrics

[Martin et al., 2012] read.

[Harvey and Thiele, 2020] read.

[Breitung and Knüppel, 2021] partially read.

3. Theory

In this chapter, we propose alternative SC-estimators to assess the magnitude of treatment effects in observational settings. To establish a general basis, we first describe the contextual environment of the estimation. Similar to the setting as introduced by ADH, we consider a framework with $J + 1$ panel units indexed by $j = 0, 1, ..., J$ that are observed over a time horizon of T periods. Without loss of generality, assume that unit $j = 0$ is exposed to the treatment at period $t = T_0$ with $1 < T_0 < T$ and that there are no treatment anticipation and contamination (i.e., no spillovers in time and space). The former would be the case if the treatment affects unit $j = 0$ before T_0, the latter describes the case where some of the supposedly untreated units $j = 1, ..., J$ are contaminated as they are affected by the treatment. To contextualize these assumptions, [Abadie et al., 2010] argue that in the presence of anticipation effects, T_0 could be shifted back in time until the no-anticipation assumption seem plausible. If panel units in the donor pool[1] are affected by the treatment (contamination) as it is likely in the Brexit-application, those units could be removed from the sample prior to the estimation. Our goal is to evaluate the causal effect of the treatment, the specific functional form of which remains unspecified though. This is possible because the main goal of the SC-estimation lies in the precise estimation of the counterfactual. Since the treatment scenario is empirically observable, it is not necessary to specify the specific functional form of the it.

The following chapter is structured as follows: We first describe the canonical estimation procedure as proposed by ADH. Furthermore, ADH propose a model-invariant hypothesis testing approach. As this approach is employed in the further analysis, we also give a brief overview of these principles. Next we build intuition by considering a simple static scenario with only two donor units and one treatment unit. This setting is subsequently generalized to the case with more potential donors. Our extensions diverge from the setting of ADH in two key aspects: First, we remove the weight constraints, leading us to explore regularization as a means to prevent overfitting. Second, we analyze a situation without covariates which drastically reduces the data requirements and causes our algorithm to estimate the counterfactual with a significantly smaller information set.

[1] To ensure direct comparability with the SC literature, we adopt most of the commonly used terms. For example, control group units are labeled as 'donors'.

However, this fact leads us to the necessity of utilizing all available information in an efficient manner and establishes our main contribution: The integration of multivariate time series approaches into the SC-algorithm. The theoretical derivation of this estimator completes the chapter.

3.1. ADH Case

We start by presenting the SC-method and the testing procedure as introduced by ADH. For the sake of comparability and due to its notational clarity, we borrow the employed notation of Abadie and colleagues. In terms of structural design, we build on the thorough presentation of the SC-method and the related hypothesis testing procedure by [Firpo and Possebom, 2018].

Setup

The estimation task can be constituted by the potential outcome framework as introduced by [Neyman, 1923] and elaborated by [Rubin, 1974]. Let $y^I_{j,t}$ be the (potential) outcome for unit j at point t in the presence of the intervention. Likewise, let $y^N_{j,t}$ be the (potential) outcome for j at point t in the absence of the intervention. ADH define the treatment effect of the intervention as

$$\delta_{j,t} = y^I_{j,t} - y^N_{j,t}$$

and introduce the indicator variable $D_{j,t}$ that takes on the value 1 if unit j is treated at period t and the value 0 otherwise. Given the assumed absence of anticipation and contamination, the following outcome is observed

$$y_{j,t} + D_{j,t}\delta_{j,t} = \begin{cases} y^N_{j,t} & (\text{if } j = 0 \text{ and } t < T_0) \text{ or } j \geq 1, \\ y^N_{j,t} + \delta_{j,t} & \text{if } j = 0 \text{ and } t \geq T_0. \end{cases}$$

The goal to estimate the causal treatment effect $(\delta_{0,T_0}, ..., \delta_{0,T})$ therefore boils down to the estimation of the counterfactuals of unit $j = 0$ in the post-treatment phase $(y_{0,T_0}, ..., y_{0,T})$, i.e. on what trajectory would unit $j = 0$ have been, was there no intervention. The basic idea of ADH is to estimate these counterfactuals as a weighted average of the donor outcomes using a data-driven approach to compute the weights. Intuitively, the weights are computed such that they optimally predict the outcomes and a set of time-invariant explanatory variables for the treatment unit in the pre-intervention phase, conditional on having a percentage interpretation. Thus, for the computation of the weights, we focus exclusively on the pre-intervention time periods $t \in \{1, 2, ..., T_0 - 1\}$. Subsequently, the counterfactuals are extrapolated by applying the calculated weights to the post-intervention time periods $t \in \{T_0, T_0 + 1, ..., T_1\}$.

Let $Y_j = (y_{j,1}, ..., y_{j,T_0-1})'$ be the vector of observed pre-intervention outcomes for unit j.[2] To distinguish treatment unit and donors, ADH collect the treatment unit in the $((T_0 - 1) \times 1)$-vector Y_0 and row-bind all donor unit vectors into the $((T_0 - 1) \times J)$-matrix Y_1. Moreover, a set of K time-invariant covariates of Y_j is observed for all panel units.[3] Therefore, let X_0 denote the $(K \times 1)$-vector of covariates for Y_0 and let X_1 denote the $(K \times J)$-matrix of explanatory variables for Y_1. To estimate the causal effect of the treatment, the SC-estimator estimates the counterfactuals $(\widehat{y}_{0,1}, ..., \widehat{y}_{0,T_0}, ..., \widehat{y}_{0,T})$ of the single treated unit for the pre- and post-intervation phase as

$$\widehat{y}_{0,t} = \sum_{j=1}^{J} \widehat{w}_j y_{j,t}^N \ \forall \ t \in \{1, ..., T\}$$

The weights $(\widehat{w}_1, ..., \widehat{w}_J)$ are constraint such that $\widehat{w_j} \geq 0 \ \forall \ j$ and $\sum_{j=1}^{J} \widehat{w_j} = 1$. It is worth noting that this constraint requires the counterfactuals to belong to the convex hull of the donors as otherwise, $\widehat{Y}_0$ will never match its true counterpart. [Abadie et al., 2010] argue that "the magnitude of discrepancy" should be calculated in advance of each SC-application. If the researcher finds that the pre-intervention values of Y_0 fall outside the convex hull of the donors, the usage of SC is not recommended. Formally, $(\widehat{w}_1, ..., \widehat{w}_J)$ is the solution of the following nested optimization problem:

$$\widehat{w}(v) = \arg\min_{w} \sum_{k=1}^{K} v_k \left(x_{0,k} - \sum_{j=1}^{J} w_j x_{j,k} \right)^2$$

with v being an arbitrary positive definite vector of dimension $(K \times 1)$ which solve the

[2] For instance, in the canonical example of [Abadie and Gardeazabal, 2003], Y_0 would be the vector of GDPs for Great Britain until the Brexit referendum.

[3] In the already mentioned Brexit-example, natural predictors of GDP are consumption, investment, government spending and net exports.

second optimization problem:

$$\widehat{v} = \arg\min_{v} \sum_{t=1}^{T_0-1} \left(y_{0,t} - \sum_{j=1}^{J} \widehat{w}_j(v) y_{j,k} \right)^2$$

Afterwards, the causal effect of the intervention $\delta_{j,t}$ can be quantified at each time point after the intervention $t \in \{T_0, T_0+1, ..., T_1\}$ as the gap between observed $(y_{0,t}^N + \delta_{j,t})$ and predicted outcome $(\widehat{y}_{0,t}^N)$.

This two-step estimation procedure serves two crucial purposes: $\widehat{v}$ measures the relative importance of the K variables in X_1 to explain X_0. In contrast, the weighting vector $\widehat{w}(v)$ quantifies the relative importance of each unit in the donor pool. Summarizing the key concept of ADH, the SC-method ensures that the synthesized treatment unit is as similar as possible to the actual treatment unit with respect to the quantity of interest and a set of potential explanatory variables in the pre-treatment period. Especially in the canonical examples of SC, the quantity of interest (e.g. GDP) and the explanatory variables (e.g. consumption, investment, government spending and net exports) are interconnected by construction. Thus, observing that the SC-estimator was capable of approximating both targets significantly enhanced the methods credibility. If the explanatory variables are omitted, the SC-algorithm reduces to an Ordinary Least Squares (OLS) estimation, constraint to have no constant and weakly positive coefficients that sum up to one.

Hypothesis Testing

The question of treatment effect significance arises naturally subsequent to the construction of the synthetic control. ADH propose a model-invariant non-parametric inference procedure that is based on the Exact Hypothesis Test proposed by [Fisher, 1935]. The basic idea behind such permutation tests is to compare the observed data with a number of randomly permuted versions of it, and to use the distribution of the test statistic calculated of the permuted samples to estimate the probability that the observed result occurred by chance alone.

In the context of SC, ADH consider permutations in region (i.e. panel unit) and time. Region permutations estimate the treatment effect vector $(\delta_{j,T_0}, ..., \delta_{j,T})$ for each panel unit $j \in \{0, ..., J\}$.[4] This procedure provides the researcher with the empirical $(J+1)$-observational distribution of the treatment. Next, it is possible to compare the estimated treatment vector $(\delta_{0,T_0}, ..., \delta_{0,T})$ of the truly treated unit with the J placebo-treatment vectors of the units of the donor pool. Given the estimated treatment effect for $j = 0$ is large, the null hypothesis of no treatment effect can be rejected at the significance level of one minus the percentile of $(\delta_{0,T_0}, ..., \delta_{0,T})$ in the empirical distribution.[5] Time permutations on the other hand consider only panel unit $j = 0$, permute T_0 to dates prior to the true treatment date and compute again the empirical treatment distribution. Given that $T_0 >> J$, this approach can increase the sensitivity of the test, since the theoretically feasible significance threshold of region permutation tests is determined by $\frac{1}{J}$. For both, region and time permutations, ADH condense the vector of estimated treatment effects into a precision metric like the Mean Squared Forecast Error (MSFE)[6] of the following form:

$$MSFE_j = \frac{\sum_{t=T_0}^{T} \left(\widehat{y}_{j,t}^{N} - y_{j,t}^{N}\right)^2}{T - T_0}$$

A possible problem that can occur when assessing the relative rarity of the estimated treatment effect using the procedure described above is the existence of outliers in the donor pool. In the context of region permutations, suppose that a donor region is very different from the rest such that it falls outside the convex hull of the remaining donors. Note, that this circumstance does not cause problems for the truly treated region and its synthesized counterfactual as we expect the SC-algorithm to assign a near zero weight to such an outlier. However, since the outlier itself cannot be synthesized precisely by the donor pool, both MSPE and MSFE are expected to be large. As this special feature causes the permutation test to be unreasonably conservative, ADH propose to exclude regions that are hard to predict, i.e. who have a MSPE that exceeds the MSPE of the truly treated unit to a great extent. Figure 1 visualizes the exclusion procedure in the tobacco control application of ADH.

[4] Note that it is necessary to exclude the truly treated unit from donor pool to ensure the validity of the no contamination assumption.

[5] For instance, let $J = 99$ such that treatment effects for 100 panel units can be computed. As long as the estimated treatment effect of the truly treated units belongs to the 95 largest effects (95th percentile or higher), the permutation test rejects the null hypothesis of no treatment effect at least at 5 percent.

[6] Note that ADH speak of the Mean Squared Prediction Error for dates before and after T_0. Since we consider the time span until T_0 as prediction window and the time span after T_0 as forecast window, we employ the label Mean Squared Prediction Error (MSPE) before T_0 and the label MSFE from T_0 onward.

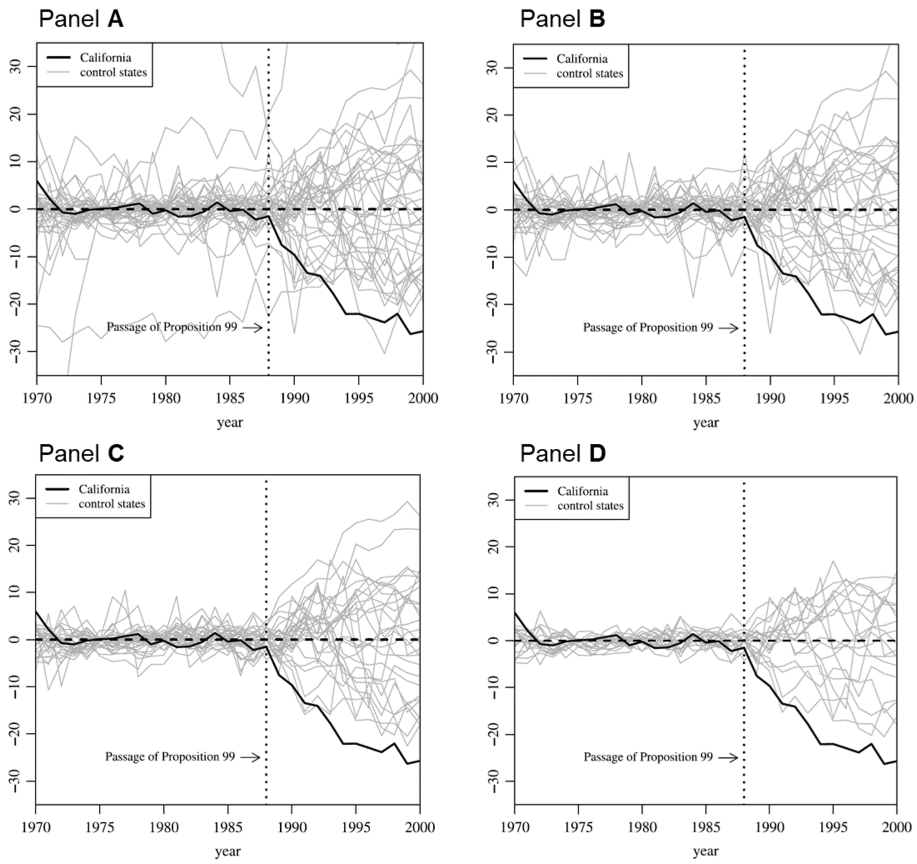

Figure 1. Region Exclusion Procedure of ADH

The vertical axis indicates the gap between observed and estimated per capita cigarette sales, the bold line represents the truly treated region (California). Two observations stand out when considering panel A: First of all, the treatment has a clear negative effect for California. Second, some regions have both a poor pre- and post-treatment fit. Since the treatment significance should not be artificially driven by regions with poor fit, ADH successively remove regions with a large MSPE relative to California. Panel B excludes regions with a MSPE that is more than 20 times as large the MSPE of California, Panel C lowers the cutoff to five times California's MSPE and Panel D to two times the MSPE. In the last scenario, only 19 regions are left and California is the one with the most extreme treatment effect. The authors therefore conclude that the treatment is statistically significant with a (permutation) p-value of 5,3% $\left(\frac{1}{19}\right)$.

One way to bypass the inefficient sample reduction procedure is to look at the distribution of the ratios of MSFE and MSPE. By scaling the post-treatment fit by the pre-treatment fit, regions with a poor fit are implicitly controlled for. In the tobacco control application, California is the region with the highest MSFE-to-MSPE ratio among all 39 regions which translates into a p-value of 2,6% $\left(\frac{1}{39}\right)$.

3.2. Simple Static Extension

To give an intuitive introduction to our proposed extensions, we first consider the most simple scenario of one treatment unit $j = 0$ and two donor units $j = 1, 2$. We consider a setting where only the outcome series (e.g. GDP) and no further covariates (e.g. consumption, investment etc.) are observed. It is assumed that before $t = T_0$ the units have a joint distribution of the form

$$Y = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix} \sim \mathcal{N}(\mu, \Sigma) \text{ for } t < T_0.$$

with $\mu = (\mu_0, \mu_1, \mu_2)'$ and the positive definite covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_0^2 & \boldsymbol{\sigma'_{12}} \\ \boldsymbol{\sigma_{12}} & \mathbf{\Sigma_2} \end{pmatrix}.$$

σ_0^2 denotes the variance of y_0, $\mathbf{\Sigma_2}$ is a (2×2) covariance matrix of the vector $(y_1, y_2)'$ and $\boldsymbol{\sigma_{12}}$ is a (2×1) vector with elements $cov(y_0, y_1)$ and $cov(y_0, y_2)$.

Disregarding any constraints, we are interested to derive the best unbiased forecast of y_0 given the controls y_1 and y_2 which is obtained as

$$\begin{aligned} \widehat{y}_0^N &= \mu_0 + w_1^{OLS}(y_1 - \mu_1) + w_2^{OLS}(y_2 - \mu_2) \\ &= \mu^* + w_1^{OLS} y_1 + w_2^{OLS} y_2 \end{aligned}$$

where $\mu^* = \mu_0 - w_1^{OLS}\mu_1 - w_2^{OLS}\mu_2$. This forecast can be directly estimated by an unrestricted OLS regression of y_0 on y_1 and y_1. However, the result implies that there is no inherent reason to impose the restrictions that $w_1^{OLS}, w_2^{OLS} \geq 0$ and $w_1^{OLS} + w_2^{OLS} = 1$. Furthermore, we argue that the construction of SC should include a constant term, as otherwise the estimated counterfactual may have a mean outside the convex hull of the donor means. See also [Doudchenko and Imbens, 2016] for a careful discussion of these restrictions.

For illustrative reasons, assume that

$$Y \sim \mathcal{N}\left(\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1 & 0.1 & 0.4\\0.1 & 1 & 0.5\\0.4 & 0.5 & 1\end{pmatrix}\right).$$

For this example the unrestricted optimal weights for the counterfactual result as $w_1^{OLS} = -0.1333$, $w_2^{OLS} = 0.4667$ and $\mu^* = \mu_0 - w_1^{OLS} \cdot \mu_1 - w_2^{OLS} \cdot \mu_2 = 0.6667$.[7] Note that w_1^{OLS} is negative even though all bivariate correlations between the units are positive. One may argue that this result does not make much sense as the economic interpretation of y_1 entering the counterfactual $\widehat{y}_0^N$ with a negative sign is unclear. This demonstrates the trade-off between optimality in a statistical sense and the economic interpretation of the solution.

What happens if we impose the restrictions that all weights are positive and sum up to unity? In this case the restricted optimum yields the linear combination $\widetilde{y}_0^N = 0.2y_1 + 0.8y_2$. The important difference lies in the variance of these estimates. For our example we obtain

$$var(y_0 - \widehat{y}_0^N) = 0.8267$$
$$var(y_0 - \widetilde{y}_0^N) = 1.1600.$$

It is interesting to note that the variance of the restricted estimate is even larger than the unconditional variance of y_0. This is possible as $(w_1, w_2) = (0, 0)$ is not included in the restricted parameter space.

So far we argued and showed illustratively, that an unrestricted OLS estimate can be superior to the constraint SC estimate in settings with few panel units and a clear correlation structure among the units. This indication will be further refined in subsequent Monte Carlo simulations. In microeconometric settings it is usually assumed that the units in the treatment group and units in the control group are uncorrelated. In such cases the

[7] The derivation of the employed estimators is postponed to the appendix.

construction of a SC is unpromising as the dependency between treatment unit and donors is the core condition for a plausible estimation of the counterfactual. In macroeconomic applications however, the variables in the treatment and control group (e.g. GDP) are typically correlated and it is therefore important to model this relationship. As the simple scenario with only two panel units in the donor pool is highly unrealistic in practice, we now move to the general static case with $J+1$ panel units.

3.3. General Static Extension

In empirical macroeconomic practice, the observed time series are typically low-frequency, i. e. the quantities of interest are measured at monthly, quarterly or even annual intervals. Thus, the number of pre-intervention time periods (T_0-1) is typically small and may even be smaller than the number of units in the donor pool J. In such scenarios, the unrestricted OLS estimate may face issues of instability or, in the case of $T_0-1<J$, due to singularity, it may not even be identified. The issues of external validity and overfitting are closely related to the aspect of identification. Especially when employing non-parametric statistical learning methods, it is simple to achieve a high in-sample (pre-treatment) fit. The crucial part when dealing with forecasts is that the observed in-sample patterns generalize well outside the verifiable horizon (post-treatment). ADH solve this issue by restricting the weights to be non-negative and to sum up to one. Besides preventing the model from overfitting, the percent restriction guarantees the existence of unique weights, especially when dealing with a small number of pre-treatment periods. Regularized regressions constitute another model family that is capable of balancing the trade-off between under- and overfitting.

ELASTIC NET

In this context [Doudchenko and Imbens, 2016] suggest using an elastic net regression to regularize the donor weights with the following objective function:

$$\widehat{w} = \arg\min_{w} \underbrace{\sum_{t=1}^{T_0-1}\left(y_{0,t}-\mu^{*}-\sum_{j=1}^{J} w_j y_{j,t}\right)^2}_{RSS} + \lambda_1 \underbrace{\left(\sum_{j=1}^{J} w_j^2\right)}_{Ridge} + \lambda_2 \underbrace{\left(\sum_{j=1}^{J} |w_j|\right)}_{Lasso}$$

The L_2-norm of the Ridge-Penalty is a continuous shrinkage method, that shrinks the coefficients towards zero without performing variable selection in the sense that certain coefficients are set exactly to zero ([Hoerl and Kennard, 1970]). However, it has the appealing feature that its estimation only involves the addition of a diagonal matrix to the Residual Sum of Squares (RSS). Therefore, the objective function keeps an explicit closed form solution which is particularly important if the sample is small.

In contrast, the L_1-norm of the Lasso as proposed by [Tibshirani, 1996] penalizes the sum of the absolute values of the coefficients. The nature of the penalty term causes this regularization to perform both, continuous shrinkage and automatic variable selection. As a consequence, the argmax of the objective function typically contains many entries that are exactly zero which makes the resulting model sparse and easier to interpret. However, since the absolute value function is not continuously differentiable, the Lasso has no closed form solution. Consequently, the minimum of the objective function has to be approximated, which is typically done via numerical optimization techniques like cyclical coordinate descent (see for example [Friedman et al., 2010]).

The shrinkage parameters λ_1 and λ_2 can be chosen by k-fold cross-validation, i. e. we store all combinations of λ_1 and λ_2 that minimize the objective function of the k hold-out pre-treatment data batches and compute the average value of the two hyperparameters.

REGSC

We propose a different regularization that we call the regularized synthetic control estimator. This estimator augments the OLS objective function by a Ridge penalty and a simple "inverse" Ridge that shrinks the coefficient sum towards one. The objective function has the following form:

$$Q(w, \lambda_1, \lambda_2) = \sum_{t=1}^{T_0-1} \left(y_{0,t} - \mu^* - \sum_{j=1}^{J} w_j y_{j,t} \right)^2 + \lambda_1 \left(\sum_{j=1}^{J} w_j^2 \right) + \lambda_2 \left(1 - \sum_{j=1}^{J} w_j \right)^2$$

This regularization is closely related to original SC estimator and in contrast to the elastic net, it is flexible enough to produce non-zero weights that are directly interpretable. Moreover, as it does not involve approximating the gradient of the absolute value function, it as a closed form solution.

than the elastic because instead of having two penalties that both shrink the parameters towards zero, we propose a weighting scheme that employs the following objective function:

The first penalty is familiar L_2-norm that conducts continuous shrinkage. The second

- argue that our regularization is more suitable in small samples due to closed from
- explain that the combination with Lasso was also programmed using numerical optimization. Argue that these methods have poor small samples properties. Cite relevant literature
- talk about Lasso and Ridge
- our goal: find regularization that is more flexible than SC but still as nice small sample properties

there are also other competing procedures to ensure weight stability and generazibility like the elastic net for synthetic controls as proposed by .

Therefore some kind of regularization is necessary to obtain a reliable estimate of $\widehat{Y}_0^N$.

3.4. Univariate Dynamic Extension

What must be clear by now

- **TBD**

When modeling macroeconomic time series it is often assumed that the $(k+1) \times 1$ vector of time series $y_t = (Y_{0t}, ..., Y_{kt})'$ can be represented by a VAR model given by

3.5. Multivariate Dynamic Extension

4. Simulation

4.1. Static Data Generating Processes

Simulation-Procedure

- [Abadie et al., 2010] suppose that the counterfactuals $Y_{1,t}^N$ for $t > T_0$ are given by a factor model of the form $Y_{i,t}^N = \alpha_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it}$. α_t denotes an unknown common factor with constant loadings across panel units. Z_i is a vector of observed panel-specific covariates, θ_t is a vector of unknown parameters, λ_t is a vector of unknown common facots and μ_i are panel-specific unknown factor loadings. The unobserved transitory shocks ϵ_{it} have zero mean at the panel level. For this specific setting, [Abadie et al., 2010] show that "[...] the bias of the SC-estimator can be bounded by a function that goes to zero as the number of pre-treatment periods increases." Further the number of donor units has to be fixed.

- [Ferman, 2021] considers a de-meaned scenario without additional covariates. The representation of the counterfactuals therefore boils down to $Y_{i,t}^N = \lambda_t \mu_i + \epsilon_{it}$. Thus the counterfactual is given by the composition of the unknown panel-specific factor loadings and the unknown common factors plus the idiosyncratic shocks.

- We re-estimated the Tobacco Control and the Basque application and found that the inclusion of additional covariates did not improve the predictive accuracy of the SC estimator. Therefore, we follow the simulation suggestion of [Ferman, 2021] and root our simulation in his proposed factor model without additional covariates.

- Analogous to Ferman, we consider a setting with two common factors, $\lambda_{1,t}$ and $\lambda_{2,t}$. The potential outcomes for the treated unit and for the first half of the donor pool load exclusively with loading one on the first factor, the remaining donors load exclusively with loading one on the second factor. Therefore μ_i is a (2×1)-rowvector with the first (second) entry being one and the second (first) entry being zero for the first (second) half of the donor pool.

- In each simulation, we simulate a total of $T_0 + T_1$ observations, where T_0 represents the length of the pre- and T_1 the length of the post-treatment period. In order to have a simulation framework that is as close as possible to real world SC-applications, we choose $T_0 \in \{20, 50, 100\}$ and $T_1 \in \{10, 20, 30\}$. Further, we define the size of the donor pool as $J \in \{5, 10, 15, 20, 25, 30\}$.

- To initialize the simulation, we simulate the two common factors by drawing $T_0 + T_1$ independent and identically distributed (iid) observations from a standard normal distribution. Next, we multiply the common factors for all donors and the treatmend unit with the factor loadings μ_i and add iid standard normal shocks. Additionally, we add panel-specific iid standard normal intercepts such that our DGP takes the following form: $Y_{i,t}^N = \gamma_i + \lambda_t \mu_i + \epsilon_{it}$.

- Figure 2 visualizes one potential DGP with $T_0 = 20$ and $T_1 = 10$. To make the factor structure visible, we scaled the factor variance by 10^1 and the shock variance by 10^{-1}. Moreover, for better eyeball inspection, we added a constant treatment effect of $\delta_{1,t} = 10$ for $t > T_0$. Note that the specific nature of the treatment effect is irrelevant for our investigation. Since $Y_{1,t}^T$ is observable for $T > T_0$, we can choose an arbitrary functional form or even implement no treatment effect at all.

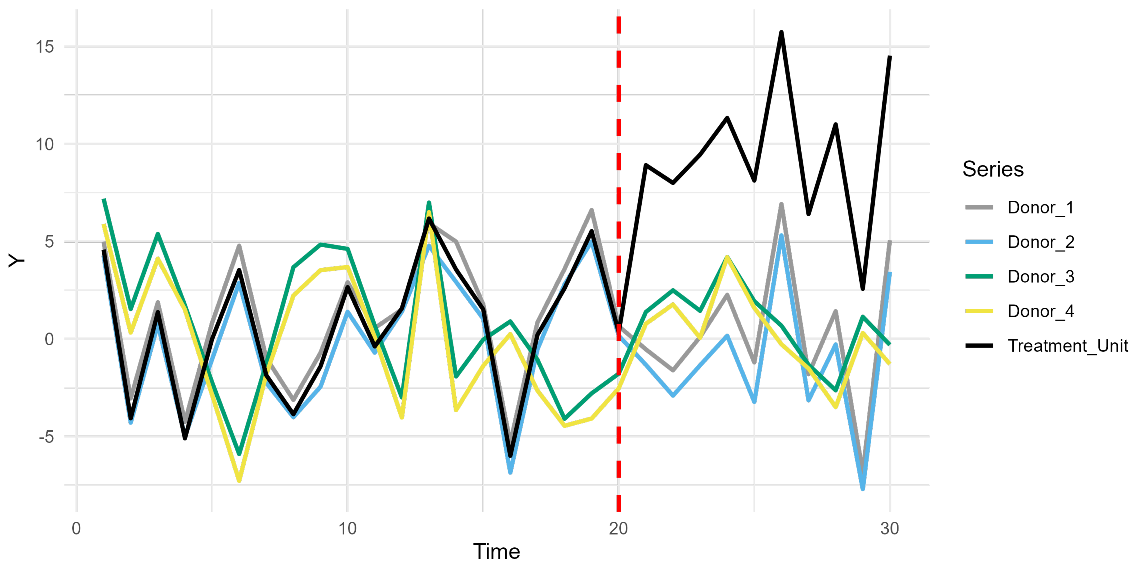

Figure 2. Example Factor-DGP

- The described factor structure is inherent in Figure 2: The treatment unit and the first half of the donors (Donor_1 and Donor_2) as well as the second half of the donors (Donor_3 and Donor_4) share a common factor structure. The constant treatment effect is added after $T_0 = 20$ periods, resulting in a 10 unit vertical shift of the treatment series.

- We consider five different models whose goal is to recover the true factor structure of the treatment unit in the pre-treatment period and to predict the counterfactual in the post-treatment period as accurately as possible. Put differently, we train the models on T_0 pre-treatment observations and test their performance using T_1 post-treatment observations.

- The following models are employed: **(Update when previous parts are written)**

  1. Factor-Model (FACTOR): It is assumed that the common factors and the idiosyncratic component are uncorrelated and the idiosyncratic errors are mutually uncorrelated. This suggests to estimate the common factors as linear function of the donors. A popular estimator with this property is the Principal Components (PC) estimator. In the training period, we obtain the predictions by regressing the treatment series on the "latent" factors. As we implemented a two-factor structure, the factors are computed by multiplying the first two eigenvalues of the covariates covariance matrix with the matrix of the covariates. The forecasts for the testing period are obtained by multiplying the factor structure of the post-treatment period with the regression coefficients of the pre-treatment regression. This model is most natural when generating data according to a factor process. Therefore, we expect it to perform best among all models.

  2. Synthetic-Control-Model (SC): In case of no additional explanatory variables, the SC-approach reduces to a constraint regression that regresses the treatment series on the donor series given the constraint of no intercept and non-negative coefficients that sum up to one. ADH argue that the constraints prevent the SC models from over-fitting, but they do so at the cost of a reduction in flexibility.

  3. Regularized OLS-Model (REGOLS): Our proposed model. We allow for a constant and arbitrary coefficient values. We propose a two-dimensional regularization: A ridge/l2-norm penalty that penalizes the coefficient sum and a

second penalty term that shrinks the coefficient sum towards one. We identify the optimal hyperparameter-combination by applying a 50/50 train-test-split in the pre-treatment period on a two-step random grid search. In the first step, we randomly select 400 hyperparameter-combinations from a (50×50)-grid. In the second step, we enclose the potential optimum by sequentially holding the first and the second hyperparameter fixed while increasing and decreasing the remaining hyperparameter on a coarser grid.

4. Elastic Net (NET): [Doudchenko and Imbens, 2016] propose to estimate the counterfactual using an elastic net regression.

5. OLS-Model (OLS): Finally, to verify the belief that a simple linear model overfits the training data and therefore yields poor forecasting accuracy in the testing period, we also employ an ordinary least squares model.

All simulations are conducted with 1,000 iterations per combination of pre- and post-treatment horizon and donor quantity.

(Important to distinguish different intercept cases: If the intercept of the treated series falls outside the donor-intercepts, SC exhibits a bias as it does not allow for a panel-specific constant.)

Table S1. Simulation Results of the Static Factor Model with $\mathbf{J} = \mathbf{5}$ Donors.

T_0	T_1	**FACTOR**	**SC**	**REGOLS**	**NET**	**OLS**
		RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]
20	30	1.2494 (0.0166) [0.6158]	1.4604 (0.0001) [1.1078]	1.2942 (0.0142) [0.6464]	1.2959 (0.0123) [0.5695]	1.3552 (0.0102) [1.1093]
	20	1.2595 (-0.0068) [0.6477]	1.4560 (0.0320) [1.0652]	1.3080 (-0.0009) [0.6610]	1.3067 (-0.0003) [0.5948]	1.3665 (0.0066) [1.0979]
	10	1.2342 (0.0250) [0.6312]	1.4322 (0.0256) [1.0473]	1.2877 (0.0256) [0.6361]	1.2865 (0.0202) [0.5516]	1.3569 (0.0309) [1.0745]
50	30	1.1819 (-0.0025) [0.6326]	1.4327 (-0.0015) [1.0579]	1.2130 (-0.0036) [0.6444]	1.1987 (-0.0014) [0.5430]	1.2133 (-0.0034) [0.7755]
	20	1.1837 (0.0092) [0.6469]	1.4224 (0.0097) [1.0563]	1.2124 (0.0144) [0.6396]	1.2046 (0.0131) [0.5583]	1.2194 (0.0150) [0.7840]
	10	1.1663 (0.0038) [0.5825]	1.3990 (-0.0019) [0.9754]	1.1966 (-0.0041) [0.5731]	1.1815 (0.0045) [0.5071]	1.1968 (-0.0014) [0.7163]
100	30	1.1636 (-0.0046) [0.6505]	1.3944 (0.0176) [1.0736]	1.1818 (-0.0060) [0.6336]	1.1746 (-0.0052) [0.5802]	1.1807 (-0.0057) [0.7192]
	20	1.1684 (-0.0009) [0.6201]	1.3973 (0.0049) [1.0416]	1.1861 (-0.0024) [0.5941]	1.1808 (-0.0020) [0.5488]	1.1862 (-0.0022) [0.6838]
	10	1.1434 (0.0127) [0.5850]	1.3686 (0.0263) [0.9784]	1.1552 (0.0114) [0.5811]	1.1528 (0.0123) [0.5151]	1.1557 (0.0129) [0.6471]

Table S2. Simulation Results of the Static Factor Model with $\mathbf{J} = \mathbf{10}$ Donors.

T_0	T_1	FACTOR	SC	REGOLS	NET	OLS
		RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]
20	30	1.1613 (-0.0051) [0.7638]	1.2867 (-0.0034) [1.0520]	1.2304 (-0.0040) [0.7474]	1.2739 (-0.0057) [0.7799]	1.6198 (-0.0090) [2.1449]
	20	1.1555 (-0.0059) [0.7316]	1.2840 (-0.0267) [1.0055]	1.2250 (-0.0059) [0.7538]	1.2595 (0.0061) [0.7655]	1.6167 (-0.0024) [2.1752]
	10	1.1312 (-0.0091) [0.6970]	1.2532 (-0.0102) [0.9745]	1.1987 (-0.0140) [0.7263]	1.2395 (-0.0149) [0.7504]	1.5811 (-0.0052) [2.0458]
50	30	1.1043 (0.0071) [0.8100]	1.2497 (0.0032) [1.0271]	1.1369 (0.0109) [0.7619]	1.1467 (0.0111) [0.6834]	1.2086 (0.0083) [1.1120]
	20	1.0966 (0.0079) [0.7854]	1.2308 (-0.0060) [1.0024]	1.1317 (0.0024) [0.7350]	1.1407 (0.0063) [0.6546]	1.2074 (0.0048) [1.0766]
	10	1.0912 (-0.0076) [0.7318]	1.2167 (-0.0053) [0.9551]	1.1204 (-0.0075) [0.7016]	1.1281 (-0.0058) [0.6186]	1.1895 (0.0001) [1.0395]
100	30	1.0851 (0.0065) [0.7875]	1.2316 (0.0110) [1.0178]	1.1063 (0.0067) [0.7389]	1.1112 (0.0080) [0.6770]	1.1323 (0.0081) [0.9248]
	20	1.0733 (-0.0137) [0.7698]	1.2088 (0.0019) [0.9747]	1.0917 (-0.0134) [0.7290]	1.0941 (-0.0125) [0.6631]	1.1202 (-0.0122) [0.9043]
	10	1.0661 (-0.0082) [0.7524]	1.1986 (-0.0119) [0.9553]	1.0895 (-0.0096) [0.7098]	1.0934 (-0.0129) [0.6451]	1.1178 (-0.0113) [0.8864]

Table S3. Simulation Results of the Static Factor Model with **J = 15** Donors.

T_0	T_1	FACTOR	SC	REGOLS	NET	OLS
		RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]
20	30	1.1355 (0.0015) [0.7829]	1.2504 (-0.0022) [1.0159]	1.2171 (0.0004) [0.7942]	1.3089 (0.0012) [1.0400]	2.4509 (0.0009) [6.0410]
	20	1.1223 (0.0082) [0.7681]	1.2368 (-0.0057) [1.0051]	1.2088 (0.0130) [0.8111]	1.3010 (0.0024) [1.1466]	2.4339 (0.0089) [6.0128]
	10	1.1199 (0.0060) [0.7029]	1.2299 (0.0152) [0.9262]	1.2142 (-0.0096) [0.7565]	1.3095 (-0.0061) [1.0792]	2.4066 (-0.0104) [5.7301]
50	30	1.0834 (0.0062) [0.8168]	1.2104 (0.0041) [0.9866]	1.1219 (0.0041) [0.7623]	1.1337 (0.0041) [0.6970]	1.2756 (0.0005) [1.3377]
	20	1.0794 (0.0047) [0.7768]	1.2016 (0.0050) [0.9709]	1.1222 (0.0043) [0.7428]	1.1324 (0.0054) [0.6599]	1.2737 (0.0025) [1.2865]
	10	1.0674 (0.0001) [0.7333]	1.1873 (-0.0077) [0.8756]	1.1010 (0.0021) [0.6879]	1.1105 (0.0026) [0.6355]	1.2532 (0.0028) [1.2290]
100	30	1.0732 (0.0036) [0.8262]	1.1780 (0.0093) [0.9729]	1.0966 (0.0043) [0.7689]	1.1032 (0.0062) [0.6935]	1.1508 (0.0059) [1.0253]
	20	1.0645 (-0.0153) [0.8304]	1.1774 (-0.0091) [0.9807]	1.0905 (-0.0185) [0.7736]	1.0963 (-0.0192) [0.7014]	1.1459 (-0.0203) [1.0300]
	10	1.0627 (0.0114) [0.7628]	1.1613 (0.0090) [0.9068]	1.0829 (0.0146) [0.7123]	1.0902 (0.0149) [0.6393]	1.1378 (0.0171) [0.9584]

Table S4. Simulation Results of the Static Factor Model with $\mathbf{J} = \mathbf{20}$ Donors.

T_0	T_1	FACTOR	SC	REGOLS	NET	OLS
		RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]
20	30	1.1278	1.2332	1.2183	1.2859	NA
		(0.0020)	(-0.0040)	(-0.0039)	(-0.0079)	NA
		[0.7964]	[1.0193]	[0.8413]	[0.9901]	NA
	20	1.1244	1.2312	1.2083	1.2712	NA
		(-0.0082)	(0.0043)	(-0.0054)	(0.0025)	NA
		[0.7614]	[0.9685]	[0.7839]	[0.9067]	NA
	10	1.1054	1.1986	1.1987	1.2632	NA
		(0.0172)	(0.0063)	(0.0171)	(0.0169)	NA
		[0.7525]	[0.9464]	[0.8319]	[0.9309]	NA
50	30	1.0703	1.1749	1.1087	1.1247	1.3707
		(-0.0056)	(-0.0160)	(-0.0070)	(-0.0088)	(-0.0067)
		[0.8408]	[0.9813]	[0.7851]	[0.7387]	[1.6546]
	20	1.0678	1.1720	1.1117	1.1237	1.3646
		(0.0207)	(0.0202)	(0.0149)	(0.0194)	(0.0054)
		[0.8154]	[0.9649]	[0.7801]	[0.7150]	[1.5848]
	10	1.0514	1.1444	1.0846	1.0976	1.3227
		(-0.0141)	(-0.0264)	(-0.0120)	(-0.0085)	(-0.0089)
		[0.7774]	[0.8997]	[0.7324]	[0.6672]	[1.4695]
100	30	1.0447	1.1422	1.0679	1.0793	1.1626
		(-0.0039)	(0.0168)	(-0.0029)	(-0.0007)	(-0.0016)
		[0.8600]	[0.9745]	[0.7904]	[0.7231]	[1.1338]
	20	1.0510	1.1476	1.0715	1.0815	1.1589
		(-0.0144)	(-0.0185)	(-0.0139)	(-0.0105)	(-0.0113)
		[0.8576]	[0.9550]	[0.7864]	[0.7234]	[1.1391]
	10	1.0304	1.1144	1.0505	1.0598	1.1397
		(0.0223)	(0.0166)	(0.0271)	(0.0270)	(0.0321)
		[0.8118]	[0.9018]	[0.7425]	[0.6826]	[1.0761]

Table S5. Simulation Results of the Static Factor Model with $\mathbf{J = 25}$ Donors.

T_0	T_1	FACTOR	SC	REGOLS	NET	OLS
		RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]
20	30	1.1093 (0.0013) [0.7989]	NA NA NA	1.2010 (0.0090) [0.8648]	1.2554 (0.0116) [0.9421]	NA NA NA
	20	1.1073 (0.0131) [0.8177]	NA NA NA	1.1989 (0.0108) [0.8516]	1.2708 (0.0093) [0.9724]	NA NA NA
	10	1.0971 (-0.0046) [0.7589]	NA NA NA	1.1928 (-0.0001) [0.8153]	1.2447 (0.0013) [0.8945]	NA NA NA
50	30	1.0597 (-0.0018) [0.8466]	1.1582 (0.0135) [0.9590]	1.0996 (-0.0048) [0.7879]	1.1172 (-0.0079) [0.7353]	1.4863 (-0.0151) [1.9825]
	20	1.0538 (-0.0150) [0.8350]	1.1525 (0.0043) [0.9332]	1.0993 (-0.0082) [0.7821]	1.1139 (-0.0101) [0.7263]	1.4811 (0.0033) [1.9856]
	10	1.0432 (0.0062) [0.8077]	1.1428 (-0.0038) [0.9111]	1.0831 (0.0062) [0.7625]	1.0997 (0.0064) [0.7214]	1.4400 (-0.0066) [1.8603]
100	30	1.0378 (0.0000) [0.8501]	1.1232 (0.0058) [0.9454]	1.0641 (-0.0003) [0.7912]	1.0741 (0.0006) [0.7149]	1.1959 (0.0044) [1.2244]
	20	1.0483 (0.0110) [0.8655]	1.1401 (0.0130) [0.9457]	1.0747 (0.0114) [0.7968]	1.0842 (0.0127) [0.7234]	1.2002 (0.0147) [1.2330]
	10	1.0296 (-0.0018) [0.8083]	1.1154 (-0.0017) [0.8879]	1.0551 (-0.0012) [0.7444]	1.0648 (0.0001) [0.6782]	1.1811 (-0.0040) [1.1463]

Table S6. Simulation Results of the Static Factor Model with $\mathbf{J} = \mathbf{30}$ Donors.

T_0	T_1	FACTOR	SC	REGOLS	NET	OLS
		RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]
20	30	1.1013 (0.0128) [0.8224]	NA NA NA	1.1899 (0.0149) [0.8865]	1.2523 (0.0115) [0.9715]	NA NA NA
	20	1.1043 (0.0014) [0.8072]	NA NA NA	1.1825 (0.0031) [0.8431]	1.2442 (0.0052) [0.9420]	NA NA NA
	10	1.0890 (-0.0075) [0.7761]	NA NA NA	1.1647 (-0.0054) [0.8107]	1.2337 (-0.0055) [0.8943]	NA NA NA
50	30	1.0622 (-0.0022) [0.8658]	1.1477 (0.0062) [0.9630]	1.1039 (0.0009) [0.8137]	1.1297 (-0.0012) [0.7852]	1.6851 (0.0059) [2.6638]
	20	1.0508 (-0.0016) [0.8700]	1.1372 (-0.0084) [0.9523]	1.0936 (-0.0009) [0.8184]	1.1190 (-0.0004) [0.7841]	1.6716 (-0.0006) [2.5964]
	10	1.0445 (0.0006) [0.8279]	1.1342 (-0.0043) [0.9218]	1.0880 (0.0052) [0.7784]	1.1061 (0.0091) [0.7648]	1.6681 (0.0174) [2.5590]
100	30	1.0420 (0.0097) [0.8864]	1.1178 (0.0172) [0.9693]	1.0662 (0.0107) [0.8258]	1.0793 (0.0094) [0.7565]	1.2394 (0.0070) [1.3694]
	20	1.0435 (0.0072) [0.8778]	1.1237 (0.0038) [0.9520]	1.0669 (0.0054) [0.8077]	1.0788 (0.0057) [0.7417]	1.2351 (0.0064) [1.3400]
	10	1.0159 (0.0027) [0.8381]	1.0890 (0.0173) [0.8876]	1.0410 (0.0041) [0.7678]	1.0465 (0.0020) [0.7076]	1.2026 (0.0038) [1.2823]

4.2. Weakly Dynamic Data Generating Processes

4.3. Dynamic Data Generating Processes

5. Simulation

5.1. Static Data Generating Processes

Simulation-Procedure

5.2. Weakly Dynamic Data Generating Processes

5.3. Dynamic Data Generating Processes

Dynamic Case:

6. Applications

We consider three leading examples:

- [Abadie and Gardeazabal, 2003]
- [Abadie et al., 2010]
- [Abadie et al., 2015]

7. Conclusion

- Some concluding remark and an outlook
- Keep short, around 1-2 pages
- Natural extension: case with explanatory variables

References

[Abadie, 2021] Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425.

[Abadie et al., 2010] Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105:493–505.

[Abadie et al., 2015] Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

[Abadie and Gardeazabal, 2003] Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93:113–132.

[Abadie and Imbens, 2002] Abadie, A. and Imbens, G. (2002). Bias-corrected matching estimators for average treatment effects. *Journal of Business and Economic Statistics*, 29.

[Abadie and Imbens, 2006] Abadie, A. and Imbens, G. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74:235–267.

[Abadie and L'Hour, 2021] Abadie, A. and L'Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116:1–34.

[Amjad et al., 2018] Amjad, M., Shah, D., and Shen, D. (2018). Robust synthetic control. *Journal of the American Statistical Association*, pages 1–51.

[Andrews, 2003] Andrews, D. W. K. (2003). End-of-Sample Instability Tests. *Econometrica*, 71(6):1661–1694.

[Arkhangelsky et al., 2021] Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.

[Athey et al., 2017] Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2017). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116.

[Athey and Imbens, 2016] Athey, S. and Imbens, G. (2016). The state of applied econometrics - causality and policy evaluation. *Journal of Economic Perspectives*, 31.

[Ben-Michael et al., 2021] Ben-Michael, E., Feller, A., and Rothstein, J. (2021). The augmented synthetic control method. *SSRN Electronic Journal*.

[Ben-Michael et al., 2021] Ben-Michael, E., Feller, A., and Rothstein, J. (2021). Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84.

[Born et al., 2019] Born, B., Müller, G. J., Schularick, M., and Sedláček, P. (2019). The Costs of Economic Nationalism: Evidence from the Brexit Experiment*. *The Economic Journal*, 129(623):2722–2744.

[Breitung and Knüppel, 2021] Breitung, J. and Knüppel, M. (2021). How far can we forecast? Statistical tests of the predictive content. *Journal of Applied Econometrics*, 36(4):369–392.

[Brodersen et al., 2015] Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2015). Inferring causal impact using bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1):247–274.

[Cattaneo et al., 2021] Cattaneo, M., Feng, Y., and Titiunik, R. (2021). Prediction intervals for synthetic control methods*. *Journal of the American Statistical Association*, 116:1–44.

[Chernozhukov et al., 2019] Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2019). Inference on average treatment effects in aggregate panel data settings. CeMMAP working papers CWP32/19, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

[Chernozhukov et al., 2021] Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021). An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls.

University of California at San Diego, Economics Working Paper Series qt90m9d66s, Department of Economics, UC San Diego.

[Cho, 2020] Cho, S.-W. S. (2020). Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: The case of Sweden. *The Econometrics Journal*, 23(3):323–344.

[Cunningham, 2021] Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

[Doudchenko and Imbens, 2016] Doudchenko, N. and Imbens, G. W. (2016). Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis. NBER Working Papers 22791, National Bureau of Economic Research, Inc.

[Ferman, 2021] Ferman, B. (2021). On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls. *Journal of the American Statistical Association*, 116(536):1764–1772.

[Firpo and Possebom, 2018] Firpo, S. and Possebom, V. (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *Journal of Causal Inference*, 6(2).

[Fisher, 1935] Fisher, R. A. (1971 [1935]). The design of experiments (9th ed.). *Macmillian.*

[Frangakis and Rubin, 2002] Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1):21–29.

[Friedman et al., 2010] Friedman, J. H., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(i01).

[Funke et al., 2020] Funke, M., Schularick, M., and Trebesch, C. (2020). Populist leaders and the economy. ECONtribute Discussion Papers Series 036, University of Bonn and University of Cologne, Germany.

[Hahn and Shi, 2017] Hahn, J. and Shi, R. (2017). Synthetic control and inference. *Econometrics*, 5(4).

[Hainmueller et al., 2011] Hainmueller, J., Diamond, A., and Abadie, A. (2011). Synth: An r package for synthetic control methods in comparative case studies. *Journal of Statistical Software*, 42.

[Hartford et al., 2017] Hartford, J., Lewis, G., Leyton-Brown, K., and Taddy, M. (2017). Deep iv: A flexible approach for counterfactual prediction. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 1414–1423. JMLR.org.

[Harvey and Thiele, 2020] Harvey, A. and Thiele, S. (2020). Cointegration and control: Assessing the impact of events using time series data. *Journal of Applied Econometrics*, 36.

[Hoerl and Kennard, 1970] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67.

[Kellogg et al., 2021] Kellogg, M., Mogstad, M., Pouliot, G. A., and Torgovitsky, A. (2021). Combining matching and synthetic control to tradeoff biases from extrapolation and interpolation. *Journal of the American Statistical Association*, 116(536):1804–1816. PMID: 35706442.

[Kuosmanen et al., 2021] Kuosmanen, T., Zhou, X., Eskelinen, J., and Malo, P. (2021). Design Flaw of the Synthetic Control Method. MPRA Paper 106328, University Library of Munich, Germany.

[Martin et al., 2012] Martin, V., Hurn, S., and Harris, D. (2012). *Econometric Modelling with Time Series: Specification, Estimation and Testing.* Themes in Modern Econometrics. Cambridge University Press.

[Muhlbach and Nielsen, 2019] Muhlbach, N. S. and Nielsen, M. S. (2019). Tree-based Synthetic Control Methods: Consequences of moving the US Embassy. Papers 1909.03968, arXiv.org.

[Neyman, 1923] Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science 5*, 4:465–472.

[Rosenbaum and Rubin, 1983] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

[Rubin, 1974] Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

[Tibshirani, 1996] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)*, 58:267–288.

[von Brzeski et al., 2015] von Brzeski, V., Taddy, M., and Draper, D. (2015). Causal inference in repeated observational studies: A case study of ebay product releases.

[Zou and Hastie, 2003] Zou, H. and Hastie, T. (2003). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

Appendix

Simple Static Extension

OLS Solution

In case the population covariance matrix is observable, the OLS-coefficients can be directly derived from it: $\left(w_1^{OLS}, w_2^{OLS}\right) = \mathbf{\Sigma_2}^{-1}\boldsymbol{\sigma_{12}}$

SC Solution

The restricted solution is can directly be derived from the covariance matrix. The first index in the square brackets indicates the row, the second the column position.

$$\begin{aligned} w_1^{SC} &= (\boldsymbol{\sigma_{12}}'[1] - \boldsymbol{\sigma_{12}}'[2] - \mathbf{\Sigma_2}[2,1] + \mathbf{\Sigma_2}[1,1]) \,/\, (\mathbf{\Sigma_2}[1,1] + \mathbf{\Sigma_2}[2,2] - 2 * \mathbf{\Sigma_2}[1,2]) \\ &= (0.1 - 0.4 - 0.5 + 1)/(1 + 1 - 2 * 0.5) = 0.2 \end{aligned}$$

$$\begin{aligned} w_2^{SC} &= (\boldsymbol{\sigma_{12}}'[2] - \boldsymbol{\sigma_{12}}'[1] - \mathbf{\Sigma_2}[1,2] + \mathbf{\Sigma_2}[2,2]) \,/\, (\mathbf{\Sigma_2}[2,2] + \mathbf{\Sigma_2}[1,1] - 2 * \mathbf{\Sigma_2}[2,1]) \\ &= (0.4 - 0.1 - 0.5 + 1)/(1 + 1 - 2 * 0.5) = 0.8 \end{aligned}$$

Variances

The variances are derived from the weights and the covariance matrix:

$$\begin{aligned} var(Y_0 - w_1Y_1 - w_2Y_2) = {} & var(Y_0) + w_1^2 \cdot var(Y_1) + w_2^2 \cdot var(Y_2) \; - \\ & 2 \cdot w_1 \cdot cov(Y_0, Y_1) - 2 \cdot w_2 \cdot cov(Y_0, Y_2) \; + \\ & 2 \cdot w_1 w_2 \cdot cov(Y_1, Y_2) \end{aligned}$$